EMBL Online Lecture Course on

Solution Scattering from Biological Macromolecules

0. General (technical questions about the course)

Q: Will the slides be available?
A: Yes, after each lecture, you will find the pdfs of the presenter's slides here: https://www.embl-hamburg.de/biosaxs/courses/embl2020/
In addition, we will send you a link to the recordings which will initially be available for a limited time (~30 days).

Q: When and how can we ask questions?
A: During the lecture, you can ask questions via the Zoom chat. Some will be answered straight away. Depending on the remaining time, some questions will then be passed on to the speaker. All questions will be summarized here. After the lectures, you can add new questions that will be answered and added to this document.

US: When joining the live event, please remember to turn off your video and audio after you are logged in. If you have no sound, select "join audio".
You will receive an email before each lecture with a new password for that session.
Any issues can be addressed to melissa.graewert@embl-hamburg.de

1. Introduction (lecture 1: Dmitri Svergun, 05.05.2020)

General

Q: What is the size range of proteins generally used for SAXS? How does it affect data interpretation?
A: All sizes of samples can be measured (from small peptides to large macromolecular complexes). The larger samples scatter stronger so the signal-to-noise ratio will be higher and you can use lower concentrated solute.

Q: Can we study proteins bound to beads using SAXS?

A: You have to take into account that the bigger beads/nanoparticles will scatter quite strongly and might “cover” the signal from your protein. But with the right strategy for data collection, yes this can be done.

Q: What are the requirements of solution composition and protein concentration range? Is there any protein concentration to size relation while collecting SAXS data?

A: At a synchrotron facility the rule of the thumb is for a 50 kDa protein you get a very good signal at a concentration of 1-2 mg/ml (more on this during lecture 3 *“Sample preparation and requirements”*). The more particles within your sample, the stronger your signal will be. But remember the golden rule: your sample should still be happy (prevent aggregation).

Q: What is the minimal amount of protein needed for SAXS?

A: This depends on the size of protein. As a rough estimation: for a typical protein 0.5 mg can already give you quite a bit of information. more on this during lecture 3.

Q: Does the presence of a reducing agent in solution (at a concentration of approx 50 mM) affect/mask the scattering signal of the protein?

A: For most chemicals at 50 mM the contrast between protein and buffer will still be sufficient so that you can still obtain a signal. However, by adding the agent you are changing the buffer composition, so you must remember to collect data from a “new” buffer that contains this agent as well (more details during lecture 3).

Q: Do metalloproteins scatter differently compared to normal proteins?

A: As the number of metals bound in such proteins are low; there are only a few atoms compared to lots of protein atoms. So they only contribute a little bit of the overall electron density. In cases such as ferritin which is a cage for many iron molecules, the situation is different and it strongly influences the scattering.

Q: Is the partial specific volume, Vp, varying a lot among proteins?

A: For proteins it does not vary too much, it is between 0.70-0.74 cm^3/g. But as it influences the contrast of the protein a lot, you should try to know it as precisely as possible if you want to use it to calculate MW. Nucleo acids: 0.53 cm^3/g.

Q: Is this method applicable to study glycans and their flexible arms which can't be studied using crystallography?

A: Yes, the effect of adding/removing glycans can be seen in SAXS. There are ways to add them to your models to improve the fit to the collected data.

Q: Can the effect of phosphorylation be studied? Does this depend on the molecular weight of the protein? Can you observe changes in R_g upon phosphorylation or change in D_{max}?

A: This depends on the degree of change that is induced. For example, if the complex opens or closes then you most likely see an increased R_g and D_{max} for the open state. If you have a structure and an idea of the change, it is good to calculate the theoretical curves (e.g. with the program CRYSOL) while planning the experiment.

Q: Given that the scattering intensity of DNA is higher than that of protein, how can this affect the measurement of obtained parameters for Protein-DNA complexes?

A: If you want to learn something about the protein, do not measure only the complex. Scattering of DNA is two times more intense than that of proteins, so your protein signal might be masked out. If you really want to see the difference of protein compared to DNA complex you should also consider neutron scattering.

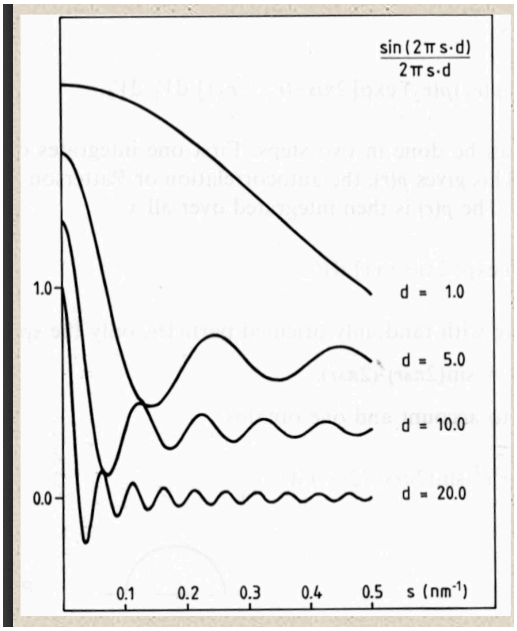
Q: Can RNA be measured in SAXS as well?

A: Yes, SAXS is done on RNA too (see lecture 8). Sometimes it can be tricky, for example, if it is prone to be digested by RNase than more precautions have to be taken.

Q: What kind of information would one obtain when performing SAS on whole bacteria?

A: Whole bacteria would scatter behind the beam stop. At very small angles you have the primary beam. Figure from slide 22, shows the fast descend in scattering intensity of large proteins. For the “large” bacteria, the descend would be even steeper. Perhaps one would better test this with light scattering.

Figure from slide 22:



Beyond the information on the whole bacteria size and outer shape, there is a limited number of studies using SAXS as a tool to get some information on the intracellular structure (e.g. membranes, chromatin,..). However, given the

complexity of a cell, the lack of monodispersity and further factors, make a straight-forward interpretation very complicated and should be taken very carefully.

Q: Is there any review paper on “how to interpret SAXS data” (that would be really nice).

A: During lectures 4+5, Al will lead you through the individual steps from data collection to interpretation. In this paper you can find the guidelines on what data/parameters to report and what to consider during the analysis/interpretation of your data: Acta Crystallogr D Struct Biol. 2017 Sep 1; 73(Pt 9): 710–728. doi: 10.1107/S2059798317011597

Q: What are Shanon channels? How do they relate to the quality of SAXS data?

A: Taking this theorem into account can sometimes make the situation more complicated. This theorem states that the amount of information you get is proportional to the number of Shannon channels that you measure. These are a discrete set of points that describes the curve (= pi/Max size, 2 pi/Max size, 3 pi/Max size), so equal spacing over the curve (in the scattering range that you record). We have programs that tell you about the amount of information (but in reality they will not tell you directly what you get) Depending on the data quality, 11 channels does this not mean you get 11 parameters. This is only a guidance.

Data processing

Guinier Analysis

Q: What sR_g limit for rod shape particle can be taken to analysed the data? The sR_g limit value for globular protein we used is $0<sR_g<1.3$, similar to this what limit we can use for rod shape particle? Somewhere I have read the sR_g limit for rod shape is 0.8 and in some paper is 1.8?

A: For rods: The overall R_g is similar to the normal limits, i.e. $s*R_g < 1 - 1.3$ within the normal **GUINIER ANALYSIS**: Plot **log(I(s))** vs **s^2** and check in this so-called **GUINIER PLOT** for the **LINEAR** regime at small s. In addition, it is possible to extract the cross-section Rg_cs. Here, one has to the perform Guinier Analysis with s*I(s), i.e. **log(s*I(s))** vs. **s^2**, and take the linear range within the curve in this Guinier plot, which can be more in the middle of the SAXS curve - in contrast to normal Guinier plot. (The reason for this difference is that the SAXS curve for a rod/elongated particle can be divided into two parts: One part at very small s, which is dominated by the long axis, and another part, at larger s, which reflect the smaller cross-section).
The most important is that it is linear in the range of the Guinier plot. Then, 0.8 is a more conservative choice. By varying the s*Rg, i.e. the s-range used to determine Rg, one can also get an idea what is the actual error of it.

Q: In order to get the radius of gyration for homodimers versus heterodimers should i do a sec-saxs experiment or is saxs alone is enough? and how can i get rid from the interference of homodimers curve in the curve of heterodimers?

A: If the separation of the two is possible with the SEC, then yes SEC-SAXS would be a good approach. If this is not feasible, it would be good to analyse the sample with other biophysical methods to understand the proportion of both. (Lecture 3 is on sample preparation. Lecture 6 on mixtures. We will go into more detail then)

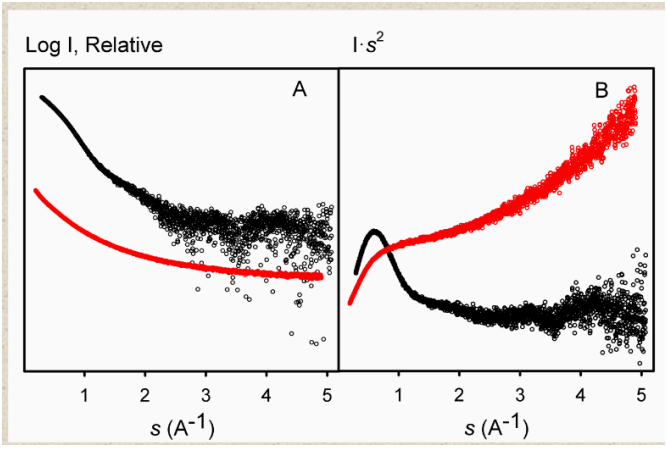
Q: In a SEC-SAXS experiment, what is the best way to determine the molecular mass? Is it possible to use the UV chromatogram to estimate a reliable concentration?

A: Besides the concentration dependent methods, there are also some concentration independent methods to determine the molecular mass, MM, from SAXS curves. Some are summarized here.
https://www.nature.com/articles/s41598-018-25355-2
For some proteins they are more accurate than others (eg. flexibility and elongation influence these estimates). For batch measurements it is good to look at I(O) derived MM estimates, as this is a good way to assess monodispersity of your sample.
Yes, you can also use UV trace to assign each frame a concentration (for this you need to know the thickness of your flow path and extinction coefficient of your protein). At our beamline and many others, we also collect light scattering data to assess MM in parallel. https://www.nature.com/articles/srep10734

Kratky plot

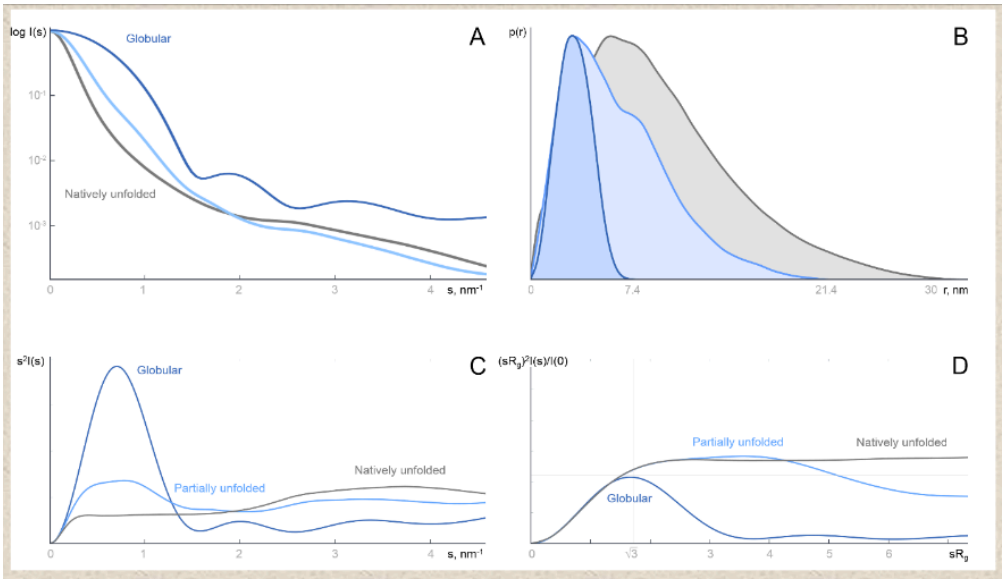
Q: About the Kratky plot on slide 38, is the red one flexible or unfolded?

A: In the Kratky plot ($I*s^2$ *vs* s, B), the bell-shaped curve in **black** is the SAXS signal for a **folded**, globular protein. The other one with plateau and linear increase at larger angles in **red** stems from an **unfolded** protein.



Q: Can I distinguish partially flexible from unfolded proteins using Kratky plot? how will it look in a Kratky plot of a protein which is folded but has a very long flexible amino terminal region?

A: Here, the middle curve is from a partially folded protein. At higher angles it will give you something going more up in the Kratky Plot. In an ideal case it would be best to compare against the same protein in its folded state. But in reality often not possible (end up running around in a circle). Best to complement Kratky analysis is other tools that address flexibility. For example our ATSAS program EOM (ensemble optimization method, see lecture 6 on mixtures). Here, one selects an ensemble of models that describe the data well. These are selected from a random pool that was for example generated by taking a compact known part of the model and modelling flexible extensions.
Short answer: always look at each case.There is no general recipe, Kratky gives you an idea but it is not inclusive to differentiate between totally unfolded and partially unfolded proteins.



Data simulated from three 60 kDa proteins: globular (dark blue), 50% unfolded (light blue) and fully disordered (gray). (A) Logarithmic plot of the scattering intensity I(s) (in arbitrary units) *vs.* s (in inverse nanometres). (B) Distance distribution functions p(r) (in arbitrary units) *vs.* r (in nanometres). (C) Kratky plot $s^2I(s)$ *vs.* s. (D) Normalized (or ‘‘dimensionless”) Kratky plot $(sR_g)^2I(s)/I(0)$ *vs.* sR_g. Adapted from Kikhney & Svergun (2015) A practical guide to SAXS of flexible and intrinsically disordered proteins. *FEBS Lett.* 589(19A), 2570-2577

Pair Distance Distribution Function

Q: Can I have an idea of the flexibility/unfolding of a protein from the distance distribution function?

A: Yes, p(r) functions resulting from unfolded/disordered proteins often show a tail with “undefined” D_{max}, see figure B above.

Q: When is big too big to measure? What's the largest D_{max} measurable with SAXS? When I use SEC-SAXS for very big proteins I don't think D_{max} is accurate (over 1000 angstrom)?

A: The minimum measured "angle" s_{min} should be less than π/D_{max}.

Q: What are the units from p(r) function?
A: If s is in nm^{-1}, r is in nm, D_{max} is in nm.

Publishing SAXS data

Q: Should all model-independent information be reported in SAXS publications?
A: The SAXS community has now introduced a Table 1 (compared to the one in crystallography). It is now required that small angle scatters report this table 1, with parameters from the experiment (some journals require this, unfortunately not all). It is also important to report experimental settings for data collection (station, temp, data range, etc) and which methods were used for analysis. The latter is often not sufficient Information such as the specific manner in which the methods were used is often lacking.
In addition it is advisable to deposit data in https://www.sasbdb.org/
See also the paper from Trewhella et al. 2017 mentioned above.
Q: I guess getting a 3D structure from a 1D scattering pattern will be a little tricky, right? Given different proteins may (I am not sure) have almost similar scattering curves?
A: There is always the issue of how ambiguous the data is. One can for example use our program AMBIMETER. One should address this when reporting on SAXS data.

2. Instrumentation (lecture 2: Clément Blanchet, 07.05.2020)

General

Q: Do you have an idea if some fee should be paid to get access to the SAXS beamline?
A: Academic users can access the P12 beamline for free (there is a proposal committee that evaluates and ranks the planned projects).

Q: Is there any minimum size limitation of the protein for SAXS?
A: You can study peptides down to 30 AAs with SAXS

Q: Can SAXS analyze polydisperse particles in solution? by ATSAS or other recommended software?
A: Yes, you get the size distributions for polydisperse systems. You can gain information about the degree of polydispersity of your sample e.g., oligomeric mixtures. This will be discussed in lecture 5 and 6.

Q: Is there a specific requirement of the type of buffers to be used or not be used for SAXS analysis?
A: In theory, the best is pure water (highest contrast). But some TRIS or salts (10-100 mM range) are ok. If your protein does not remain happy then you cannot collect suitable data. When possible do not use heavy additives, they scatter and reduce contrast. This will be discussed in lecture 3.

Q: Is there a tool that we can use to check the quality of data and if the obtained parameters are correct?
A: In SAXS, there is nothing like R-free. The quality of the fit is given by Chi-squared or Cormap p-value (in ATSAS package). More in lecture 5.

Q: Is it possible to measure samples in a 96 well setting?
A: Yes, at P12 we have a holder for 96 well plates.

Q: Why is the SAXS curve going down and has many kinks?
A: The curve decreases because of the properties of the Fourier transform (see the first lecture). Kinks you see only for roundish samples, not anisometric. The SAXS curve you refer to is from apoferritin. It is similar to a hollow-sphere - and this produces a very pronounced scattering pattern. Such minima always point to very defined symmetries.

Hardware

Q: Does the user have to select different settings while collecting data?
A: For typical proteins, you can use standard settings. In some cases, you might change the distance of the detector or choose a different wavelength/energy to look for example at larger particles. This depends on which beamline you are using. At P12 one setting is usually sufficient to get proper angular range.

Q: How much exposure is good for home source SAXS beam? for a 40 kDa protein? Does exposure depend on concentration used?
A: This depends on the home source, in general, around a few minutes. The higher concentration the better statistics at the constant exposure time.®®

Q: What material is used for focusing X-rays?
A: In case of mirrors - it is fused silica or ceramic glass compound with low coefficient of thermal expansion (such as Zerodur®) covered with a thin layer of metal.

Q: In regard to the X-ray focusing lenses, are these lenses doped? (In case of neutrons there are doped with MgF_2 etc.).
A: There is no need for doping the lens material. One utilizes the fact that vacuum has a higher index of refraction than matter - for x-rays. Therefore, fabricating defined cavities in a low Z material is a way to make X-ray lenses. Of course,

the reflection power is rather low, hence more lenses are usually employed in a row.

Q: How do you decide what array of lenses to use for focusing at a particular distance (how to vary the number of lenses)?
A: For X-ray lenses, one uses low Z elements. Classically Berylium, aluminium lithium, as these have low absorption, as the X-ray beam passes through.

Q For small protein, big protein, virus - do you use different detector distances?
A: In principle, it is a good idea to change the distance, but at P12 the middle distance is good for everything - the detector is large and thus a large s-range is covered.

Q: What are the main differences between the detectors of SANS and SAXS? If there is any?
A: Neutron detectors are based on a completely different design as these rely on the interaction of neutrons with matter to produce a detectable signal. Also, the pixels are much larger. SANS detectors are still largely gas-wire based.

Special sample environments

TR-SAXS

Q: If a protein is showing a conformational change in the presence of some ligand is it possible to study it using time-resolved (TR) SAXS?
A: Yes, conformational changes can be studied by TR SAXS.

Q: In the Laser Triggering for TR SAXS, can we use the caged proton and change pH in solution?
A: Yes. Caged-compounds are indeed the perfect target for laser-induced TR-SAXS

Q: Can we in-situ UV irradiate the sample while scanning the sample using SAXS?
A: Technically this is possible but not at P12 at the moment.

Temperature scans:

Q: How would one perform temperature-dependent SAXS studies of proteins? The data collection time for a protein at a synchrotron ranges from 1s to max 60s. So in very less time how to achieve a higher temperature to analyse the temperature-dependent effect.
A: With our advanced set-up we can go down to a lot faster exposure times (for example during the time-resolved experiments). For temperature changes that do not occur rapidly one can heat the sample holder while loading (with automated sample changer) and exposing aliquots at various temperature steps along the temperature ramp. For more sophisticated temperature studies there are dedicated sample environments that get placed into the beamline.

Q: Is it feasible to collect data at cold temperatures?
A: Cold temperatures are possible, but it is more a question of sample stability in case of biological SAXS. Lowest feasible temperature is around 5°C with our standard setup and 0C with our dedicated temperature stage. A cryo-sample environment can also be installed on the beamline where samples can be cooled down at liquid nitrogen temperature, but precise temperature control is not possible (and operation is a bit cumbersome, contact us if interested).

SEC-SAXS

Q: What flow rate is used for SEC-SAXS?
A: The flow rate depends on the column specs. Normally we flow at 0.3-0.5 ml/min (if we go slower, we often end up "burning" the sample). Therefore not all columns are practical for SEC-SAXS (more on this in lecture 6).

Q: How to collect data to minimize buffer background with SEC-SAXS?
A: For SEC-SAXS experiments we use frames that are collected before or after the peak and contain scattering of just the buffer for background subtraction. To help with the processing of this step we use the program CHROMIXS (more during lecture 6).

Q: Is the MALS at P12, a 3-row detector or 18-row?
A: The MALLS set-up at P12 measures laser light scattering at 3 different angles.

3. Sample and buffer preparation (lecture 3: Melissa Gräwert, 12.05.2020)

General

Q: Can non-Europeans **access the beamline**?
A: At EMBL beamlines, users outside Europe are welcomed. Here, you will find the link to apply for beamtime and further information. https://www.embl-hamburg.de/biosaxs/user_info.html

Q: Do the beamline scientists **assist with data analyses** and structure determination if users without experience apply?
A: At our beamline we have collaborative mode for inexperienced users. This can be requested during the beamtime application procedure.

Q: How can one **differentiate between different proteins** in SAXS? can one see single residues or „only“ secondary structure?

A: With SAXS one cannot distinguish between single residues. In some cases in which single point mutations lead to conformational changes (eg. open vs. closed states) then this can be distinguished. In general, different proteins usually produce different scattering patterns. One can calculate the scattering from an atomistic model and compare it to the experimental SAXS data (eg. Program Crysol, lecture 4).

Q: In the case of protein mixtures is it better to use SANS instead of SAXS?
A: SANS gives similar info to SAXS for mixtures. Thus for a mixture of proteins SAXS is a better method. In case of DNA/protein complexes it is better to use SANS.
SANS makes sense only for complexes where one part of the complex can be “made invisible” for neutrons.

Q: Does it make sense to quantify the amount of the large and the small particles in the sample by the area of the peak in SEC analysis? (if there are two kinds of particles in the sample, one is large, and the other is small).
A: Yes, absolutely. The more you know about the sample the easier to interpret/validate your SAXS analysis. Mixtures will be discussed in lecture 6.

Q: Can we collect the SAXS data for only RNA and how to prepare the RNA sample for analysis? While trying to optimize the buffer for RNA, what would you suggest to check/play with? Currently what I see is the longer the RNA is in the FPLC buffer, the more it aggregates?
A: Yes, there are quite a few examples of SAXS and RNA. In preparation for collecting SAXS data on RNA it is best to communicate with your local beamline responsible in advance and discuss what precautions you normally use for your sample preparation and see if they can be applied at the beamline as well. For example we have RNase ZAP available to prevent digestion of RNA through RNase contamination. There is some flexibility in the FPLC buffer you can use. Perhaps find a solution with less salt, so that your sample still elutes but stays “happy”.

Q: Can I measure the membrane protein with detergent and determine the protein's model? For membrane proteins if we subtract the buffer it won't account for detergents bound to proteins? how to subtraction so that we can get normalize for scattering from ordered detergent molecules upon addition of protein to buffer as in buffer alone detergents won't be ordered

A: Membrane proteins are usually modelled in SAXS including the surrounding micelles or detergent belts. This can be done for example with modelling two phases. Here, SANS may be helpful. SEC-SAXS is a good approach to remove free micelles directly before data collection.

Sample requirements:

Q: Does high concentration affect the structure factor?
A: Yes, so there is a limit for the concentration. Therefore, doing a concentration series is often a good approach.

Q: Can we use different gradient of concentrations for data collection? if yes what are the ideal gradients?
A: Typically, something like 0.5, 1.0, 2.0, 4.0, 8.0 mg/ml is sufficient.

Q: What about multicomplex components: when we use various concentrations, we perturb the equilibrium of binding with each other?
A: If the Kd is high, i.e., low affinity, then dilution will cause the complex to disassociate. This is especially important to consider for SEC-SAXS or SEC-SANS where the column dilutes the samples. Complexes need a low micro- to nano molar Kd for SEC. Otherwise use batch SAXS to keep the concentration high, e.g., at a concentration 5-10 Kd.

Buffer/ additives:

Q: Why does the signal increase in low q, is it sample aggregation or buffer problem?
A: This can be both aggregation and buffer mismatch.

Q: Is there a better buffer for SAXS experiments? For example PBS, HEPES or Tris?
A: The buffer choice is highly dependent on the protein. Some buffers result in higher radiation damage though. Often samples in PBS often tend to show more effects of radiation damage then in HEPES. TRIS is also frequently used (in SEC mode it sometimes shows a weird drift).

Q: How long should we perform dialysis?
A: If you know the sample survives (does not aggregate over time), typically overnight, at least 2-3 changes of dialysis buffer. At minimum 8 hrs with 3

buffer changes. You also have to take surface area of the dialysis membrane and beaker volume into account.

Q: What concentration is recommended for DTT and glycerol? For Batch and for SEC-SAXS?

A: DTT = up to 10 mM, typically 5 mM. Make sure your protein is happy in a reducing environment! Glycerol = up tp 5% v/v...higher concentrations compromise the intensity signal for SAXS.

Q: Are there better/worse salts that can be used? Is it the matter of salt concentration, or should I try adding other compounds?

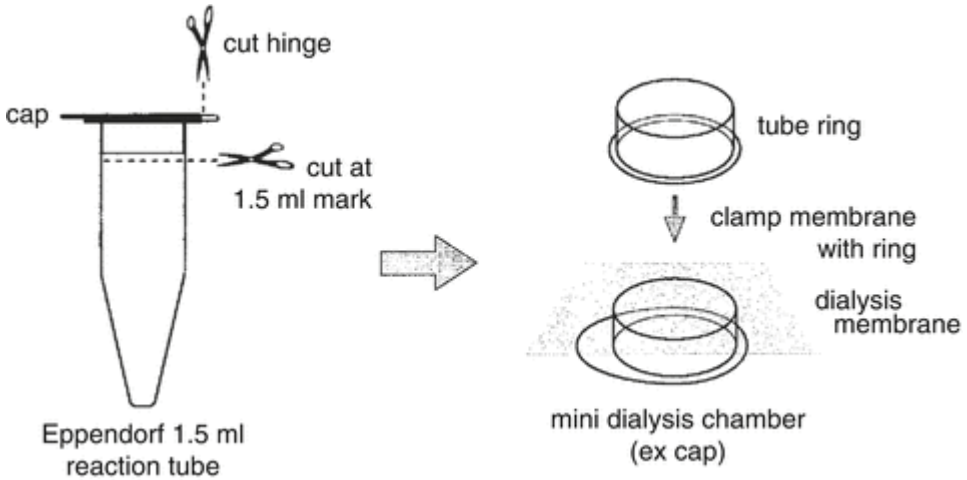
A: The best is to reduce the amount of additives as much as possible. However, keeping the sample in a “happy” state should be the main aim. Different buffers and additives can be screened for example as described here: https://www.embl-hamburg.de/services/spc/Visitors/Service_list/screen_1/

Q: Which reducing agent do you suggest for SAXS measurements?

A: Usually DTT or TCEP for intracellular proteins, ascorbic acid for extracellular proteins. Remember that DTT does not have a long shelf life. Thus, if possible add freshly to the sample/buffer before the measurement.

Q: Can you share details about the home-made dialysis device you mentioned?

A: Here is a picture taken from https://link.springer.com/protocol/10.1007/978-1-4939-2474-5_1



You can fill up to 200 ul into the ex cap. For harvesting the sample after dialysis remove excess volume for the top of the membrane with a tissue paper and puncture a whole with the pipette tip and aspirate the dialysed sample.

Sample storage/ shipping

Q: What can one do if storage at high concentrations of sample leads to aggregation or precipitation?

A: Upon storage, one should filter (eg. with 0.2um spin filter) or centrifuge at high speed to precipitate the aggregates.

Q: What is the best way to ship/bring samples for SAXS analysis to the synchrotron? What is better, 4 degree or stored in liquid nitrogen?

A: Theoretically, it is better NOT to freeze the samples -- might create additional problems but practically proteins are not stable for too long at higher temperatures. You should test beforehand the stability at 4C for the time needed between preparation and measurements, or the stability upon freeze/thaw. This can for example be checked with DLS or SEC in advance in your lab.
If the sample does not aggregate on defrosting dry ice is recommended. Otherwise send on blue ice (approx. 4C). Always inform the staff at the facility how to store the samples on arrival.

Analysis

Q: Can we combine data from both SAXS and WAXS to generate a high resolution data?

A: Yes, it is possible. For example with the program PRIMUS you can merge two profiles. This will be explained in Lecture 5.

Q: How can I determine the theoretical P(r) to calculate Dmax in different conformational states?

A: You can calculate the theoretical scattering curve from your models (for example with CRYSOL). You can then apply indirect Fourier transform to these scattering profiles using GNOM from the ATSAS package (Lecture 4).

Q: If you have models representing different states of the protein, you can fit your SAXS data with a weighted mixture of these models.

A: If you have models of these states one can try to fit the data by a linear combination of theoretical SAXS curves (OLIGOMER from ATSAS). This will be covered in Session 6.

4. Data processing and *ab initio* analysis (lecture 4: Al Kikhney, 14.05.2020)

General

Q: Choosing the exposure time is one of the key factors. So my question is: how do we fix the exposure time ? or it is a sample specific phenomenon that varies along with type of sample. Please explain.

A: At P12 we usually collect many frames with a short exposure time. If they are different, that is most probably the radiation damage. it is better to have many frames at low exposure times and average identical ones.

Q: How to decide the fraction for Guinier plot? How small of s^2 is near I(0)?

A: One normally inspects the linearity of the Guinier plot, and checks a range such that $s_{max} * R_g < 1.3$.

Q: How to calculate the partial specific volume of the buffer component?

A: You can use the Program MULCH:

http://smb-research.smb.usyd.edu.au/NCVWeb/

MULCh: Modules for the analysis of small-angle neutron contrast variation data from bio-molecular assemblies.Whitten, A. E., S. Cai, and J. Trewhella (2008)

Q: How to generate paper quality figures using ATSAS?

A: In primusqt, you can personalize the appearance of the plots. For this go to Plot >Configure in the menu. Then you can save the current view as xxx.png.

Q: Is it possible to open SAXS data with EXCEL, *SciDAVis* or Origin?

A: Reduced SAXS data are normally .dat files in ASCII format, with three columns for s, I, error, these can be directly imported in excel or origin.

Concentration dependencies and data merging

Q: Can a shift in scattering pattern at different concentrations also suggest that radiation damage occurred?

A: If there is a shift of profiles at different concentrations, it is good to check each profile for radiation damage. If at least the first two frames are the same then you can assume that you collected the first data frames before the onset of radiation damage. If you suspect radiation damage and have enough sample, you could try to collect with even lower exposure times. Concentration effects due to different oligomeric states are similar at different exposure times

Q: When merging different concentrations. What is the region in s that we need to align?

A: To merge profiles make sure there are enough points that overlap between the two curves and align these overlapping points (you might need to scale along the y-axis). For choosing which curves to merge a guide may be: "for the low angle portion pick the highest concentration which does overlap well with the lowest one"

Q: How many points can we cut while merging the data?

A: Inspection of a concentration series thoroughly. Normally at the lowest concentration there should not be much effect on the low-angle portion so that you do not have to remove many points.

Q: Are you always able to merge the data?

A: Ideally, you will have performed enough dilution steps and still have high signal to find good curves to merge. Sometimes you can not merge if there is not enough overlap between two curves. If the “low” concentration curve has a good enough signal-to-noise ratio up until the s range that you want for your analysis then you can continue with this file without merging.

Q: How to choose the data to merge in SEC-SAXS?

A: For SEC-SAXS you typically do not have interactions between the molecules (diluted sample) so you can just average over the peak .

Q: If we collect different SEC-SAXS data for protein-DNA complexes changing the P:DNA ratio, we can merge the data?

A: Better not, unless you are sure the peak corresponds always to the same stoichiometry

Molecular Weight estimates

Q: How much error in molecular weight calculation is allowed using I(0) values?

A: We normally assume 5-10% error. But if the MW is off, this is a good indication to recheck your concentration measurement or re-examine polydispersity of your sample (eg. with SEC, DLS).

Q: How much error in molecular weight calculation is allowed using I(0) values?

A: In practice, given all uncertainties in concentration, partial specific volume etc. the accuracy of MW is about 5-10%. Assessment from Porod volume is empirical and not too accurate.

Q: Can different approaches give different MW? Differences of 50%? Does this mean the data is not good?

A: Indeed different methods can give slightly different values, here one can utilize the Bayesian approach: Hajizadeh et al. (2018) *Scientific Reports* 8:7204

50% off is too much. In practice given all uncertainties in conc, partial specific volume etc accuracy of MW is about 5-10%. Assessment from Porod volume is empirical and not too accurate.

Q: Is there a software for the Bayesian approach?

A: The program DATMW in the ATSAS package. It is also included as a tool in ATSAS primus. It gives you a consensus estimate based on Bayesian analysis of multiple estimates and seems to be quite reasonable. Generally, different methods work differently for different shapes. The credibility interval estimated by DATMW can tell you more about the accuracy of mw estimation.

Q: How do you determine MW on absolute scale for SEC-SAXS? (is normalising vs concentration just merging the curves around the peak?

A: You will have to estimate the concentration for the individual frames for example through calculation from the UV trace (remember to use the right Extinction coefficient). At P12, we offer RI and MALLS with our SEC-SAXS set-up. If you have data sets without this information then it is OK to report SAXS data without I(0) derived MW. The reason I(0) derived MW is given much weight in the Trewhella et al. 2017 paper, is that it is a good way to check sample quality of batch data and show that if this value is as expected your sample is indeed monodisperse. With SEC SAXS you are normally producing a monodisperse sample, so if you don't have the MW from I(0) you can still use the data. (That is, if you don't have capillary fouling).

Q: Can flexibility of your protein affect the results from any of the techniques for MW estimation?

A: Theoretically flexibility does not matter because all electrons still scatter in phase. Practically there may be effects like increased hydration but they are minor

Q: For us, we use SEC-SAXS with no standard. The automatic analysis still somehow calculates Porod Volume but it's never close to MALLS molecular weight - should we be ignoring porod volume values?

A: In SEC-SAXS you unfortunately can often get capillary fouling thus your data has impurities (despite the separation step). If Porod Volume is too high, make sure you don't have such a fouling effect or that you have the best buffer for subtraction.

Calculating the p(r) function

Q: What are the likely problems if D_{max} doesn't tend to zero?

A: It is either aggregated or small angles are not small enough because the molecule/particle is too big for the current measurement settings

Q: For D_{max} it is sometimes possible to fit data with 112, 113 and 114 nm; without the crystal structure how can we say which is best or which value we have to report?

A: A difference between 112 and 113 in D_{max} is negligible. If it were dmax values between 8,9 and 10 nm then take a better look at the fits.

Q: What is a good agreement between I(0) and R_g between the real and reciprocal space, given that these are calculated using different parts of the curve. Is an error of 2-5% reasonable and is there a cutoff? say >10% means poor analysis?

A: Even 5% error is too much -- if you eg use GNOM the errors are minimal

Q: Can a longer tail in D_{max} indicate unfolding/flexible N terminal from your sample?

A: A long tail is usually an artifact pointing to partial aggregation of your protein.

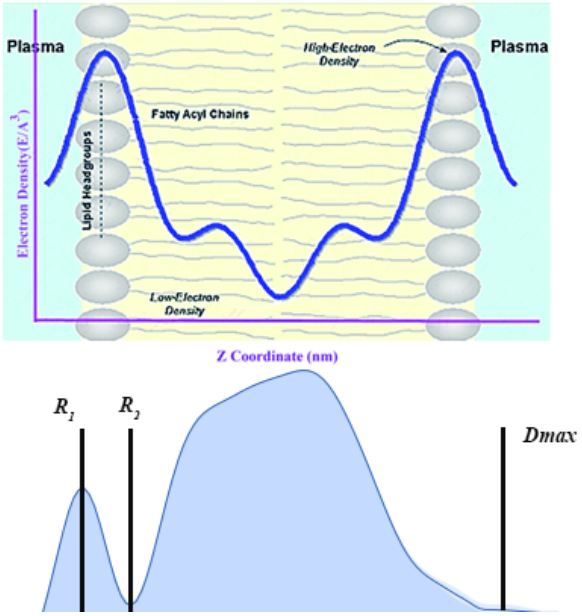
Q: Is it also possible to directly open ATSAS p(r) file in excel?

A: No. But you can open it in a text editor and copy the p(r) at the end of the xxx.out file into your spreadsheet. With command line one can use: $out2pofr xxx.out -o file.pddf.

Q: Can I use ATSAS software package to calculate the distance distribution function from lipids bilayers? I want to check if I get accurate information about any abnormalities (size fluctuations) in the bilayer structure when we add certain drugs or particles from SAXS data, by comparing data in real space.

A: You can use the same program GNOM. At the small r (central scattering part up to r = R1) the p(r) corresponds to the distances within the core of the lipid structure, and the standard SAXS analysis is applicable. When lipids are present, the main difference in terms of scattering are regions of various electron density and contrast. As the electron density of hydrophobic regions is lower than those of proteins (and may be even lower than the electron density of the surrounding solvent), the p(r) function can reach negative values at its minimum at r = R2 (see figure below). Here, the R1 and R2 correspond to the size of the inner part (if you e.g. have some particle within a lipid micelle) and the size of the hydrophilic region, respectively. The maximum dimension of the overall particle can be found the same way as usual from p(r) and Dmax. Therefore, even without any further modeling, the characteristic sizes of your system can be readily estimated directly from pair-distance distribution function and the R1, R2, and Dmax values. Please note that the solution is assumed to be monodisperse and particles have identical R1, R2, and Dmax; otherwise you will get some averaged model over the ensemble of all particles. If you are interested in the further modeling, a program DAMEMB coupled with MONSA (both from the ATSAS

package) may be used to ab initio reconstruct the model of your system in 3-phase approximation (particle, hydrophobic tails, hydrophilic heads).



Ab initio modelling

Q: If we know the symmetry of our molecule from crystal structure, should we still always use P1 while ab initio modelling?

A: Do both P1 and symmetry reconstruction and compare -- it is useful

Q: Can you specify symmetry such as D3 or D2?

A: If you are used to using Schoenflies notation for point groups: for D2 use the P22 (or, equivalently, P222), for D3 the P32 option.

Q: What parameters would you first consider to be altered, when doing ab initio (symmetry, alpha, refining etc. ?)

A: If ab initio models do not fit your data, then go back and look at your p(r) function and make sure that fits the data. For Symmetry always start with P1 and then try other symmetries. Run AMBIMETER to check how ambiguous your solution will be.

Q: Why are the envelopes produced by DAMMIN or DAMMIF not the same in different runs?

A: DAMMIN starts always with different random models. Both programs perform a Monte-Carlo type search.

Q: When presenting *ab initio* models, I should present the averaged or the representative model? (the averaged shape doesn't fit the experimental data though)

A: For DAMMIN/DAMMIF, you should run DAMAVER on the reconstructions, which will be aligned and averaged. This runs also a program called DAMSEL, which scores the reconstructions and allows you to get the most representative one.

Q: What is the main difference between DAMMIN/DAMMIF and GASBOR? I get great envelopes with GASBOR, but less nice with DAMMIN/DAMMIF (my protein has a hole in the middle which GASBOR picks out but DAMMIN/IF doesn't?

A: GASBOR uses higher resolution data and may give higher resolution models. If the protein is not bigger than 660 kDa, give GASBOR a try.

Q: Should *ab initio* models produced by DAMMIN/DAMMIF/GASBOR be similar? Which one of the models should I use? Should all of these models be aligned?

A: For rigid proteins giving consistently overlapping reconstructions within each method, the outcome should have a similar overall shape, ie. if ambiguity of your model is not too high, they should look similar

Q: Can we use *ab initio* modelling to reconstruct Protein-DNA complexes? What if the recommended program for ab initio modelling of protein-DNA complexes when the protein structure is available?

A: For a protein-DNA complex you can use MONSA, where the protein part is fixed -- there are options for doing this.

Q: It would be very useful if we get some more information on multiphase modelling using MONSA and how to analyse data using that

Q: How do we validate the model generated by MONSA is correct or not if we do not have any structure of our complex? Is there any plot by which we could look into that?

A: For MONSA, run multiple times and check reproducibility

Q: Can *ab initio* modelling be conducted with flexible proteins?

A: Yes, but if there is extensive flexibility the solutions will be unstable (large NSD)

Q: Which program I shall use to model the intrinsically disordered proteins

A: Flexible proteins: you better use dedicated programs (will be presented later in lecture 6). Shape determination typically just gives something anisometric without much meaning

Q: Is AMBIMETER usable with extended molecules or IDP? or with these extreme cases EOM would be better?

A: Better always use programs like EOM for flexible systems. Will be discussed in lecture 6.

5. Hybrid rigid body modelling - (Al Kikhney - May 19)

General

Q: All these Data available online used in these examples?

A: Example data is provided in the help folder of the ATSAS package. Al listed the SASBDB entry codes so that you can access the data from there https://www.sasbdb.orge.

Q: Where to ask questions when running into problems

A: You may use the Saxier Forum, small angle scatterers from around the world can give you answers: https://www.saxier.org/forum/

Q: Is there a difference in the on-line versions of the programs?

A: ATSAS online versions of all programs have limitations in the number of parameters you may vary, so they are of course more convenient but more restricted.

Fitting parameters

Q: What will be the value to check for the alignment with SUPCOMB?

A: NSD - normalized spatial discrepancy.

Q: Should Chi square always be near to 0?

A: No for a good fit it should be near 1. The deviation depends on the number of points. But one should be careful with error estimations. Wrong error estimations can give a lower chi^2 even for a bad fit. The chi^2 values depend on the reliability of error estimates, which are not always accurate. In practice 1.5 is typically OK.

Q: What does it mean if my Chi^2 is below 1?

A: Chi^2 below 1 points to wrong experimental errors - probably the errors are overestimated. You can't improve errors (unless you redo the 2D→1D radial averaging and further data reduction). One can then try cormap /datcmp to check (these do not use errors for the estimate), see Franke et al. (2015) *Nat. Methods* 12(5):419-422.

Q: Any other guidelines you “improve” Chi^2?

A: You should always check that artifacts at very small or very larger angles are removed before running Crysol. In particular, any possible contribution of the direct beam, improper masking etc should be not used for calculating chi^2.

One note about chi^2: Sometimes eye is better than chi.

Take a look at the fit. See where there is agreement or disagreement. In low s range → then you might have higher species present. At wide angles, maybe you have buffer issues.

Q: Two shapes have similar fit with the data and approximately similar chi square values then which shape we should prefer?

A: You can check the overall ambiguity of your data (Ambimeter). You can also cluster your models with damaver and damsel to see which features the models share (lecture 5).

SASREF

Q: Does SASREF apply to DNA-protein complexes too?

A: Yes SASREF can be used for protein-DNA complexes

Q: Does SASREF apply to single protein with multidomains?

Q: Yes you can run SASREF on a single protein with multiple domains. Upload structures of the domains; you can "connect" these by defining the contacts between the C-terminal of the first domain and the N-terminal of the next domain. If some linkers are missing you can then use BUNCH or CORAL. Here the missing residues are “mimicked” as beads.

Q: How can the SASREF output models be validated?

A: Assuming they fit the SAXS data and are interconnected and do not have clashes, the thing we do is looking at the interfaces -- whether they are meaningful. Check if you have any biochemical hints to support (point mutations, cross linking etc). Run SASREF with and without constraints and see how robust the answer is. Also re-run SASREF to see whether the result is repeatable.

Specific cases:

Q: I have structures calculated by DAMMIF, averaged/filtered with DAMAVER programs and refined with DAMMIN with chi^2 between 1-1.1. What is the best way to fit/dock high res. structures within these envelopes? Would this be in SASREF?

A: DAMSUP should overlay high resolution structure to the bead model. also SUPCOMB can superimpose / align an atomistic model to a damaver or dammin model

Q: How do I run CORAL for a single domain dimer. Do we provide monomer while using P2 symmetry or we have to provide dimer?

A: You provide the monomer and use Symmetry: P2.

Q: I have a dimeric protein with a missing loop on each monomer in a crystal structure. Problem is that the missing part is not identical on each monomer in the xtal structure (missing 21 aa in chain A and 33 aa in chain B and conformations are different). Which program is the best in this case? I tried all CORAL, BUNCH and EOM, and tried to fix the parts that I knew from the xtal structure but it did not work. If I let one part be flexible then the final model is not what I expected (in comparison to the xtal structure).

A: Confirm that you really have a dimer (eg. first make ab initio modelling to check whether you do have a dimer). If yes, then CORAL should work - use it in P1 and fix the known parts. If your assumption is incorrect CORAL will try to stretch out the added parts.

Q: Which software is recommended to merge multi-domain pdb file when I try to build a full-length atomic model with multi-truncated domain?

A: To "merge multidomains" if this is a single chain then use BUNCH or CORAL

Q: I modelled my data with CORAL (with P1). However, the model does not fit quite well with *ab-initio* model though. So the most important is the fitting with the experimental data using CRYSOL? and I can conclude which model is better based on chi-square values?

A: You may use SUPCOMB to align your atomistic model with *ab initio* “envelopes”. If you for example used CORAL to add a missing N-terminal, then this might be flexible.

Q: If I had a protein complex with two proteins, one of them have available crystal structure, the other one has two domains and the crystal structure of two domains are available respectively, what kind of modeling program should I use?

A: This is a very good example to use SASREF. Upload the three crystal structures. Add a “contact” between the two subdomains of the 2nd protein. You also can use CORAL if the linker is longer.

6. SAXS studies on mixtures, assemblies, flexible systems (Melissa Gräwert - Tuesday May 26)

General

Q: Would it be possible to download a copy of the recordings if it is permitted?

A: Unfortunately we are not allowed to deposit a downloadable version, only streaming is available. We will collect feedback at the end of this series and decide how we will proceed from there.

Q: I am modelling a multi-domain protein complex. All the domains are globular in shape. The final model fits well in SAXS data, how can we validate the orientation of each domain?

A: You may re-run modelling a few times. Also compare with any biochemical data available. Cross linking, mutation studies. You could rearrange the individual domains with Saspy and see how the fits change.

SEC-SAXS

Q: What size range are we talking about for SEC-SAXS?

A: The size limit is based on the column you use. There are special columns for small peptides but also columns for large complexes in the MDa regime.

Q: So we don't need to collect different concentrations in SEC-SAXS because different fractions mean different concentrations?

A: The intensity along the peak is proportional to the concentration but normally a peak corresponds to a sufficiently dilute sample, which means there should be no detectable structure factor effects. The dilution of your sample happens naturally on the column.

Oligomer

Q: How were you able to determine the weight of each species on the scattering curves (referring to Insulin study)?

A: The MW was not determined in this study. OLIGOMER was used for mixture analysis. I.e. the known models were used to fit these data and find the volume ratios. We determine therefore the volume (or weight) fractions.

Q: I have a similar case, but instead of analysing the contribution of different oligomeric states, I would like to analyse the contribution from two (very different) dimeric conformations. (I have high resolution structures for both dimers, but I would like to include both simultaneously in the analysis and

determine the contribution of each one in equilibrium instead of analysing vs theoretical scattering from each model separately in crysol). Can I use oligomer for this too?

A: Yes, you can use OLIGOMER for that.

Q: What analysis tool would you use for transient dimer? Which data if data was collected from different concentrations? OLIGOMER?

A: For a transient dimer you may use GASBORMX or SASREFMX

Q: Oligomer: what if you have PDB from homologous dimer only and your protein is a mixture of dimer and tetramers (AUC and SEC-SAXS)? Which model for the tetramer?

A: Here you can try SASREFMX: You can define the homologous dimer as P1 and let the program build P2 - taking dissociation fractions into account. So it will find a dimer-dimer interface (resulting in your tetramer) but also take some of the original dimer (starting building block) into account. Here the volume fractions are determined in the same way as in OLIGOMER.

EOM

Q: Does EOM need all-atoms models? like generated from your X-ray data + modeller to fill in missing residues and loops?

A: EOM input is flexible. You may upload all-atoms structures, or substructures. Based on the sequence input, EOM checks the pdbs to see what is available and what parts to 'simulate' as beads.

Q: Is it possible to use EOM on nucleic acids, specifically DNA? I imagine that only the sequence file would be different but I have not gotten it to work using the web server version.

A: If one has a pool of randomized DNA models, one may give them as input for EOM, this will work. However, ATSAS does not have an internal program for DNA models generation (like RANCH for proteins). We did not yet have requests for EOM on nucleic acids - in principle, such an extension could be possible.

Q: Where can we learn more about how to use EOM?

A: You can take a look at the slides here. These are from a whole lecture

given by Dmitri on the analysis of flexible proteins.

7. Beyond standard BioSAXS - Thursday May 28:

Anomalous small angle X-ray scattering (ASAXS) - Andrey Gruzinov

BioSAXS - special applications - Martin Schroer

ASAXS

Q: It was mentioned that for ASAXS, CHOOCH are based on MAD phasing, but I don't understand because I thought there's no phase in SAXS.

A: The modifications in f, i.e. f', is a Phase contribution. So, the basics of scattering (in particular phase problem) are always there for (non-coherent) X-ray scattering. CHOOCH is a general usage program that allows one to calculate f' and f'' based on fluorescence measurement.

Q: Is protein concentration accuracy an issue in ASAXS?

A: Protein concentration accuracy shouldn't be a problem in ASAXS. The contribution of the anomalous part is independent from the concentration. However, with higher concentration you get the better signal-to-noise ratio. For ASAXS, the relative amount of anomalous scattering atoms and their contribution are the critical part. One ideally needs the sufficient number of "bound" or "fixed", i.e. not floating atoms. Otherwise the influence of the background scattering becomes the problem.

Q: ASAXS is only to check the binding of proteins to metals? Is it possible if I modify a small molecule compound with Br and measure ASAXS to check its binding to my protein?

A: The absorption edge should be at reasonable wavelength such that you can measure good SAXS. In principle Br is one of the good elements. ASAXS is used to get Information how the anomalous atoms are distributed. In this case you could check "binding" with normal SAXS. However. In theory it could be used if the Br is pointing outside or at the binding interface (however you likely better have more than one Br to get a measurable signal).

Q: The compound in my case binds to the protein and changes the scattering curves of both dimer and monomer but I still do not know where the binding site would be. I wonder if it is worth to try ASAXS after I modify the ligand

with Br.

A: In principle yes but note that you need two labels and the anomalous signal is weak so it is better to have double Br in each. One can also do several runs of SASREF to get an idea about possible binding sites if you have crystal structures of the components and SAXS curve from the complex.

Q: I have a very naive and basic question, how does Br label on our protein? By element replacement?

A: Brominated residues or isomorphous replacement eg Cl-Br or Ca to Tb.

Q: Would it be useful for [2Fe2S] cluster containing proteins? I have an apo-structure and I reconstitute the Iron-sulfur, so I know from Raman that the [2Fe2S] forms, but could I use ASAXS to check about conformational changes before and after the reconstitution?

A: Yes, you may detect distribution of Fe; S is more difficult. One can try to get the information about relative distance distribution of the Fe from ASAXS, although it is recommended to do the modeling with CRYSOL beforehand.

Q: These kinds of BioSAXS signals can be only detected by batch, right?

A: The ASAXS is of course only for batch, for SEC-SAXS one would need liters of sample. However, one can run the SEC-SAXS at different X-Ray energies. Due to the natural dilution of the sample in the column this can dramatically decrease signal-to-noise ratio that is even more crucial for ASAXS.

Time-resolved SAXS

Q: If some azo dye which is photoresponsive, changes from trans to cis state by UV irradiation and the life time of cis state is in the range of seconds. Will one be able to measure the cis state using this technique?

A: Yes, there are studies with SAXS to study cis/trans sites of proteins. For this the users brought their own set up and turned off all lights for dark state

and then triggered the transition.

Besides performing the whole experiment in the dark, the radiation damage for this specific sample has to be checked in particular, as a) the sample might not be flown to ensure that exactly the UV-exposed region of the sample is probed by the X-ray beam and b) it might be more susceptible to the X-ray beam itself.

Q: But is this possible to measure the cis state if it is stable just for a few seconds?

A: Yes. SAXS measurement as such can be done in millisecond range. In our normal measuring mode, we can collect frames of 30-50 milliseconds, so yes you could collect enough frames during the "activated" state. One can also set up the stopped flow for the measurement of the transition within seconds. With the EIGER 4M detector, data collections of 1.3 ms are possible. However, for a few seconds time constant, the Pilatus 6M with 35 ms time exposure per frame is sufficient.

Q: Can we validate radiation damage and temperature-induced aggregation in time-resolved temperature-changed SAXS?

A: Beamline pipeline compares the frames and discards those that have radiation damage. So in principle it is possible. Of course one should do it carefully to be sure that aggregation is temperature induced and not rely only on automatic data processing.

In this specific experiment, several approaches seem feasible (likely even more):

A) If the temperature induced aggregation is rather slow, and needs temperatures < 40 °C, at P12, the sample changer operation can work: keep sample at temperature T at storage and load at a given time a part of sample to collect SAXS patterns.

B) If higher temperatures are needed and moderate time-scales, at P12 use the in-air temperature capillary holder, and load several capillaries with sample. Keeping these at the relevant temperature, collection at different

times SAXS patterns from different batches can be done.

C) If the sample scatters strongly, you can put an absorber in the beam to attenuate the dose.

D) Repeating the measurements but starting the collection at different delays (pump probe).

It really depends on the type of samples, the time scales and the temperature range.

8. SAXS and nucleic acids - June 2nd

Q: Can you use this software to look at dna-rna interactions as well?
A: RNA Masonry models only RNA at the moment. protein-RNA complexes are also not yet supported.

Q: When using MONSA for protein-DNA complex: should we use two scattering data (1) protein alone and (2) protein-DNA complex?
A: Yes. Or three: protein alone, dna alone, complex of course the assumption is: the conformation alone and in complex is the same.

Q: Just out of curiosity: is the RNA pseudo-knot an index of flexibility?
A: Pseudoknots in RNA stabilize tertiary fold. Their formation, however, may depend on environmental factors, thus be related to a flexibility of a molecule.